DOCUMENT RESUME

ED 344 894 TM 018 173

AUTHOR Case, Susan M.
TITLE Validity of NBME Parts I and II for the Selection of Residents: The Case of Orthopaedic Surgery.
PUB DATE 92
NOTE 6p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Admission Criteria; *College Entrance Examinations; Comparative Testing; *Graduate Medical Students; Higher Education; Physicians; *Predictive Validity; Scores; Statistical Significance; Surgery; *Test Use; Test Validity; Test Wiseness
IDENTIFIERS *National Board of Medical Examiners; *Orthopedic Certification Examination; Orthopedics

ABSTRACT

The predictive validity of scores on the National Board of Medical Examiners (NBME) Part I and Part II examinations for the selection of residents in orthopaedic surgery was investigated. Use of NBME scores has been criticized because of the time lag between taking Part I and entering residency and because Part I content is not directly linked to knowledge and skills required in residency. NBME Part I scores were obtained for 481 of the 1,050 examinees who took the written component of the certification examination of the American Board of Orthopaedic Surgery (ABOS) in July 1988. Scores on Part II were available for 461 of these examinees. Statistically significant relationships were found between the ABOS examination and both NBME examinations. Part II of the NBME was a better predictor of ABOS performance than was Part I of the NBME. This study supports the belief that those who have done well on examinations continue to do well, possibly because of good test-taking skills. While the NBME examination should not be the sole determinant of acceptance in a residency program, the degree of correlation suggests that intelligent use of these scores provides an efficient and effective way to screen large numbers of applicants. One figure illustrates the relationship between failure on the ABOS and NBME scores. There is a 17-item list of references. (SLD)

VALIDITY OF NBME PARTS I AND II FOR THE SELECTION OF RESIDENTS: THE CASE OF ORTHOPAEDIC SURGERY

Susan M. Case, Ph.D.
National Board of Medical Examiners

Validity of NBME Parts I and II for the Selection of Residents: the Case of Orthopaedic Surgery

Susan M. Case, Ph.D.
National Board of Medical Examiners

Objectives

The purpose of this study was to investigate the predictive validity of NBME scores for the selection of residents in Orthopaedic Surgery.

Perspective

In selecting the best applicants for their programs, residency program directors face a considerable challenge. Many programs have hundreds of applicants for each available position, and it is common practice to use scores on the National Board of Medical Examiners (NBME) Part I and II examinations to identify applicants for further consideration (Nungester, 1990; Wagoner, 1986; McCollister, 1988). This use of NBME scores has been widely criticized, particularly for Part I, because of the time lag between taking Part I and entering residency, and because the content of Part I is not directly linked to the knowledge and skills required in residency. The NBME has also criticized this use of NBME scores because the examinations are not designed for this purpose (Volle, 1988).

A first step in assessing the appropriateness of this use of NBME scores is to determine the strength of the relationship between the predictor variables (NBME scores) and the criterion (some measure of success in residency). Typical studies have used as criterion variables either in-training exams with small sample sizes from a single program, showing inconsistent results (Spellacy, 1985; Warrick, 1986, Catalano, 1989); or ratings of resident performance, showing consistently low positive relationships (Keck, 1979; Markert, 1989; George, 1989; Yindra, 1988; Turner, 1987; Veloski, 1990; Distlehorst, 1988; Williams, 1987, Gunzberger, 1987; Wood, 1990).

In contrast to previous research, this study used scores on a professionally developed specialty board certification examination as the criterion measure. While some believe that a relationship between NBME scores and other multiple choice tests would reflect only method variance, the use of certification exams as a criterion has some basis in practicality. Program directors clearly want to accept applicants who will succeed in their programs, and one measure of success in residency is subsequent performance on the specialty board certification exam.

Data Source and Method

Subjects. A total of 481 of the 1050 examinees who took the written component of the certification examination of the American Board of Orthopaedic Surgery (ABOS) in July 1988 was identified in the NBME data base by self-reported social security numbers. These individuals took one or both of the NBME Part I or II examinations. NBME scores were obtained for 481 examinees on Part I and 461 examinees on Part II.

Instrumentation. The ABOS examination was administered in a single site under secure conditions. The six-hour examination included 274 multiple-choice questions (MCQs). The examination assessed application of knowledge through use of clinical vignettes combined with 100 radiographs, histosections, or other pictorial material that required examinees to interpret the information and formulate a diagnosis or a management plan.

The NBME Part I exam contained approximately 980 MCQs, covering the basic bio-medical sciences of anatomy, behavioral sciences, biochemistry, microbiology, pathology, pharmacology, and physiology in approximately equal proportions. The NBME Part II exam contained approximately 900 questions covering the clinical sciences of internal medicine, obstetrics and gynecology, pediatrics, preventive medicine and public health, psychiatry, and

surgery in approximately equal proportions. Part I is typically taken in the second year of medical school; Part II is typically taken in the fourth year.

Results

Characteristics of the Sample. Because only 481 of the 1050 ABOS examinees could be identified in the NBME data base, the representativeness of the sample was investigated by comparing ABOS exam performance of the sample; the ABOS reference group (ie, graduates of LCME accredited schools taking the exam for the first time); and the ABOS total group. The mean percent correct score of the sample (X=73; SD=5) was slightly below that of the reference group (X=75; SD=5), and slightly above that of the total group (X=70; SD=5).

Performance on NBME Exams. Mean scores for the sample were 514 on Part I (SD=92) and 494 on Part II (SD=91). Mean scores for those who passed the ABOS exam (n=429) were approximately 1 SD higher than for those who failed the ABOS exam (n=52) on the NBME total scores and all subscores, with the exception of behavioral science, where the difference was only 22 points.

Correlations between scores. Statistically significant relationships were found between the ABOS exam and all NBME Part exams. The ABOS exam score had an observed correlation of 0.49 with NBME Part I and 0.55 with Part II. The strongest correlations with Part I subtests were with physiology, anatomy, and biochemistry (.48, .44, and .41, respectively); somewhat weaker relationships were found with pharmacology, microbiology, and pathology (.39, .39, and .37, respectively); the weakest correlation was with behavioral science (r=.18). The strongest correlations with Part II subtests were with medicine and surgery (.50 and .48, respectively), followed by pediatrics and obstetrics/gynecology (.44 and .42, respectively); the weakest relationships were with preventive medicine and public health and with psychiatry (.37 and .36, respectively).

Predictions of ABOS scores. Regression analyses indicated that Part II was a better predictor of performance on the ABOS exam than Part I (R^2 of 0.30 vs 0.23). Using Part I subscores as predictors was only slightly better than using the total Part I score (adjusted R^2 of 0.26 vs 0.23); using subscores for Part II did not result in a higher R^2 than using the total score. Using all Part I and II subscores was slightly better than using both total scores (adjusted R^2 of 0.33 vs 0.31).

Figure 1 shows the likelihood of failure on the ABOS exam as a function of performance on the Part I and II examinations. For example, of the 19 examinees who scored below 350 on the NBME Part I exam, 11 (58%) failed the ABOS exam; of the 25 who scored between 350 and 400, 9 (36%) failed the ABOS exam. Similar results were found for Part II. The standard errors associated with failure rates are relatively large for scores under 350 (ie, approximately 10), but decrease in the remaining sections of the curve.

Educational or Scientific Importance of the Study

This study supports the belief that those who have done well on exams in the past continue to do well on exams, but there are at least three potential explanations for this phenomenon that merit discussion. The first explanation, which is endorsed by critics of MCQ exams, is that these correlations are largely a reflection of test-taking skills, not knowledge. This argument is not as compelling for certification exams taken by physicians as it is for tests in elementary and high school. These examinees have demonstrated their ability to take tests, and these tests are more carefully crafted than standard classroom tests; the item flaws that reward testwiseness are virtually non-existent in these certification exams.

The second potential explanation for the strong relationships found in this study is that performance on these tests indirectly reflects general ability, motivation, study skills, and other general traits that influence learning. Past achievement may be a good predictor of future achievement because of this indirect assessment.

If either the first or second explanations told the whole story, the correlations between NBME subscores and the ABOS score would have been uniformly high. However, the pattern of correlations found in this study suggests a third explanation: scores provide direct information about an examinee's knowledge in "job-related" sub-areas of interest. The correlations between the ABOS exam scores and Part I and II subscores range from 0.18 to 0.50, and the pattern of correlations is quite interpretable. Very little of the content covered by Part I is directly measured on the ABOS exam; however, much of it, especially in areas such as anatomy and physiology, is necessary as a framework for the knowledge assessed on the ABOS exam. As expected, the relationships with these two subjects (plus those in medicine and surgery) are relatively strong, whereas the relationships with behavioral science, psychiatry, and preventive medicine and public health are relatively weak.

The correlations between scores on the ABOS and NBME exams (taken as many as nine years earlier) are probably even stronger than they appear because the examinee group that took the ABOS exam was a select group: some of those who performed poorly on NBME Part I failed to graduate from medical school; many failed to get into an orthopaedic residency program; others failed to complete the orthopaedic program. At each step, some of the lowest performers dropped out, leaving a more homogeneous group of examinees who took the ABOS certification examination. This homogeneity attentuates the relationships that would have been observed in an *applicant* population.

There are several reasons why these results must be interpreted with caution. First, the study involved only a single specialty exam; results may not generalize to other specialties. Second, the study included only a subset of those who took the ABOS exam; others were not found in the NBME data base, because some did not take NBME exams, and because of errors in self-recording social security numbers. Replications with additional administrations of this specialty board and other specialty boards are planned for the future.

No one believes that NBME Part I and II exams should be the sole determinant of acceptance in a residency program; clearly other factors should be considered and the relationship is less than perfect: some examinees with very low NBME scores passed the ABOS exam. On the other hand, the study showed that Part I and II scores strongly predict performance on a certification exam taken many years later Those who scored below 500 on Part II were 20 times more likely to fail the ABOS exam than those who scored above 500; dramatic relationships were found between failure rates on the ABOS exam and performance on both Part I and Part II at various other cut-points. Intelligent use of these scores appears to provide an efficient and effective method for providing a preliminary screen of applicants, especially for programs with large numbers of applicants.

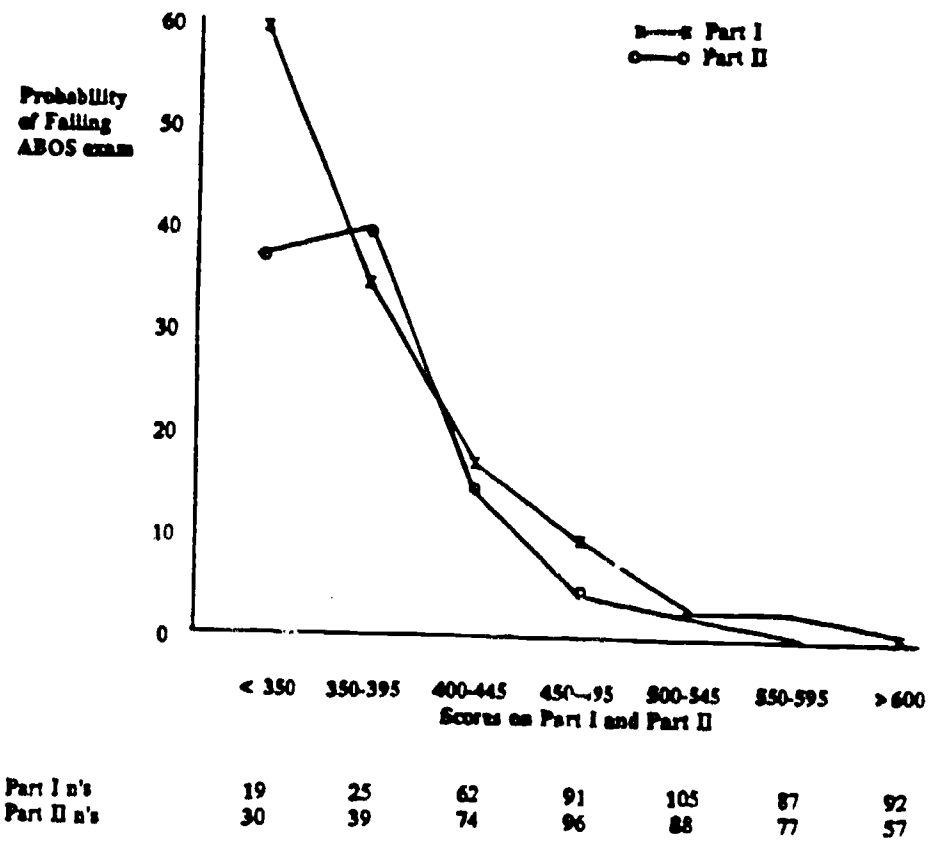

Figure 1. The relationship between failure rates on the ABOS exam and NBME Part I and Part II scores.

References

Catalano, R.A. & Jenkins, P.L. Correlating Factors to Resident Performance in an Ophthalmology Program. *Ophthalmology*, 96:552-554, 1989.

Distlehorst, L.H., Verhulst, S.J., Vu, N.V., & Barrows, M.D. The Use of Performance-Based Assessment Scores and Traditional Measures in Predicting First-Year Residency Performance. *Research in Medical Education, Proceedings of the Twenty-seventh Annual Conference* (1988):176-181.

George, J.M., Young, D., & Metz, E.N. Evaluating Selected Internship Candidates and Their Subsequent Performances. *Academic Medicine*, 64:480-482, 1989.

Gunzburger, L.K., Frazier, R.G., Yang, L., Rainey, M.L. & Wronski, T. Premedical and Medical School Performance In Predicting First-Year Residency Performance. *Journal of Medical Education*, 62:379-384, 1987.

Keck, J.W., Arnold, L., Willoughby, L. & Calkins, V. Efficacy of Cognitive/Noncognitive Measures in Predicting Resident-Physician Performance. *Journal of Medical Education*, 54:759-765, 1979.

Markert, R.J. (1989, August). Cognitive Predictors of Residency Performance. *Resident and Staff Physician*, pp. 53-58.

McCollister, R.J. The Use of Part I National Board Scores in the Selection of Residents in Ophthalmology and Otolaryngology. *JAMA*, 259:240-242, 1988.

Nungester, R.J., Dawson-Saunders, B., Kelley, P.R., & Volle, R.L. Score Reporting on NBME Examinations. *Academic Medicine*, 65:723-729, 1990.

Spellacy, W.N. The Use of National Board Scores in Selecting Residents for Obstetrics and Gynecology. *American Journal of Obstetrics Gynecology* , 153:605-607, 1985.

Turner, B., Hojat, M., & Gonnella, J. Using Ratings of Resident Competence to Evaluate NBME Examination Passing Standards. *Journal of Medical Education*, 62, 572-581, 1987.

Veloski, J. J., Hojat, M., & Gonnella, J.S. A Validity Study of Part III of the National Board Examination. *Evaluation and the Health Professions*, 13:227-240, 1990.

Volle, R.L. Using National Board of Medical Examiners Scores in Selection of Residents. *JAMA*, 259:266, 1988.

Wagoner, N.E., Suriano, J.R., & Stoner, J.A. Factors Used by Program Directors to Select Residents. *Journal of Medical Education*, 61:10-21, 1986.

Warrick, S.S., & Crumrine, R.S. Predictors of Success in an Anesthesiology Residency. *Journal of Medical Education*, 61:591-595, 1986.

Williams, R.G., & Colliver, J.A. A Study of Errors Attributable to Use of NBME Part I Scores for Residency Selection Purposes. *Research in Medical Education, Proceedings of the Twenty-sixth Annual Conference* (1987):43-47.

Wood, P.S., Smith, W.L. Altmaier, E. M., Tarico, V.S., & Franken, E.A. A Prospective Study of Cognitive and Noncognitive Selection Criteria as Predictors of Resident Performance. *Investigative Radiology*, 25:855-859, 1990.

Yindra, K.J., Rosenfeld, P.S. & Donnelly, M.B. Medical School Achievements as Predictors of Residency Performance. *Journal of Medical Education*, 63;356-363, 1988.